Naini H Mistry. Looking Beyond the Eye: Pupillary Dilation and its Relationship to Conversational Behavior in Computer-mediated Conversation and Face-to-Face Conversation. A Master's Paper for the M.S. in I.S degree. April, 2005. 23 pages. Advisor: Barbara M. Wildemuth

Research has demonstrated that pupil dilation can be an effective indicator of human cognition. When two people converse face-to-face, the body language, gaze, eye movements or visual cues in general play an instrumental role in determining the course of the conversation. However, it is not known how our pupils might actually react in a computer mediated conversation.

Nine subjects participated in a study conducted at the Interaction Design Lab at School of Information and Library Science, University of North Carolina at Chapel Hill. The study was divided into two phases. The first phase consisted of a face-to-face conversation and the second phase consisted of an online chat conversation. An eye tracker was used to obtain pupil dilation measurements in both phases. The graphs generated by the gaze tracker were then compared. Overall, mean pupil dilation was greater in the face-to-face conversation than in the online chat conversation. Implications of this finding for both system design and future research are discussed.

Headings:

Pupil dilation

Eye tracking study

Computer mediated conversation

LOOKING BEYOND THE EYE:
PUPILLARY DILATION AND ITS RELATIONSHIP TO CONVERSATIONAL BEHAVIOR IN COMPUTER-MEDIATED CONVERSATION AND FACE-TO-FACE CONVERSATION

by
Naini H Mistry

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirementsg
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2005

Approved by

____

Barbara M. Wildemuth

INTRODUCTION:

When we speak about conversation we usually imply interaction with another person. This interaction could take place through different mediums which are usually defined as communication channels. With the advent of computers and machines, interaction is no longer seen as simply a human affair but extends to humans and machines and humans and humans through machines. Researchers and developers strive to improve the usability of interfaces to make them more user-friendly, such that the user understands and is able to interpret the behavior of the machine. There have been different theories; such as the connectionist theory, that categorize the design of the computers similar to the human brain consisting of networks and connections that are capable of processing information in parallel. Does this imply that, when humans interact with computers, they envision them as being human as well? The emulation approach to attaining human-computer collaboration is based on the metaphor that we endow computers with "human-like abilities."

Pupil dilation has long been correlated with aspects of human behavior. When we hear about sayings such as "eyes are windows to the soul," it essentially means that eyes are capable of communicating our mental activities. Theorists suggest that the pupil dilates in response to interest, cognitive load, anxiety and personality, in addition to physical responses to light.

In a face-to-face conversation, visual feedback often plays an instrumental role in determining the course of the conversation. Factors such as eye movements, turn-taking, hand gestures and body language in general, function as effective cues in gauging the role of listener and speaker. These cues also imply that our eyes are also constantly communicating our thought processes during the course of the conversation.

In the case of an interaction with computers, there is again a sense of a dialogue that occurs between the user and the computer. When humans interact with humans through the medium of computers it is often difficult to determine the process of turn-taking since visual feedback affects our actions before speech. Chat interfaces offer features that are metaphors of human emotions, but is the communication value the same? It is often difficult to maintain the recipient's attention in a conversation that occurs via the computer. So how does a user visualize the interface while chatting with another user? Does the same amount of dilation occur as it would have, in a normal face-to-face conversation? If so, are we treating computers as humans when we interact with them? Why is there such a struggle to create more and more attention grabbing user interfaces? Is eye-contact that important during an interaction?

Studying the pupil behavior in these two environments, viz. face-to-face and online chat, can lead us to an understanding of the importance of eye contact during a conversation, especially in applications such as online or distance-based education or conferencing tools.

LITERATURE REVIEW:

Human computer interaction is a multidisciplinary field that studies how people interact with computers. Interaction implies active communication between two entities. Computers have gone through a metamorphosis from being mere devices that carry out calculations to agents that are capable of carrying out complex tasks allocated by the user. These tasks are facilitated by the interface which, in turn, receives the commands the user inputs and provides the feedback the user interprets. This is indeed comparable to a two-way communication that an average human being has with his/her counterpart. However, there are obvious human factors, such as senses, emotions etc., which are significantly absent in the case of computers.

One fundamental factor that distinguishes interaction between humans from that with computers, as Vertigaal et al. (2000) have described, is conversational cues such as gestures, facial expressions, looks and tone of voice. These cues not only determine the meaning of the words we use, but also play an essential role in regulating the conversational process.

According to Short et al. (1976), as many as eight cues may be used: completion of a grammatical clause; a socio-centric expression such as ‘you know’; a drawl on the final syllable; a shift in pitch at the end of the clause; a drop in loudness; termination of gestures; relaxation of body position and the resumption of eye contact with a listener. According to a recent study by Vertigaal and Ding (2002), 49% of the reason why someone speaks may be explained by the amount of eye contact with an interlocutor.

Another important aspect that governs conversation is the process of turn taking. We often look for visual feedback in our conversations to determine who can next lead the conversation. As Vertigaal and Ding (2002) pointed out, humans use eye contact in the turn taking process for three reasons:

a) Eye fixations provide the most reliable indication of the target of a person’s attention, including their conversational attention.

b) The perception of eye contact increases arousal, which aids in proper allocation of brain resources, and in regulating interpersonal relationships.

c) Eye contact is a *nonverbal visual* signal, which can be used to negotiate turns during turn-taking.

We can relate these findings to human and computer interaction. Maglio et al. (2000) found that, in human-to-device verbal communication, people look at the devices with which they wish to communicate before issuing any spoken commands. They based their observation on the theory of looking before speaking in human-to-human conversation. In their experiment they had participants interact with an Attentive User Interface and it was noted that the participants defaulted to a natural social condition of conversation with the devices as though they were human.

These studies showcase the importance of visual feedback in a dialogue between entities. At the same time they do not undermine the existence of audio applications or digital personalities that facilitate conversation as well.

We often hear the proverb that eyes are the mirror/mirrors of the soul. Time and again books illustrate that eyes are windows to a person's thoughts. John Senders remarks that eyes speak the mind; where he says,

"The eyes are the window to the mind and the mind's window to the scene
So that one is never quite sure whether it's the world or the mind
That makes the eyes shift to where they're going from where they've been"

- *John W. Senders, "Visual Scanning Processes" (1983)*

Knapp and Hall (1997) noted that, by observing the eye we can obtain reasonable inferences about human emotions. Previous research has demonstrated that the pupil of the eye is associated with other human behaviors. This could be cognition, mental effort, or emotional factors such as anxiety, interest, novelty, etc.

For centuries, people have taken measures to either enhance or diminish the "communication" value of the pupil. For example, as mentioned in Janisse (1977), Gump(1962) has reported that the wise Occidental customer would wear dark glasses when conducting business with an Oriental jade merchant, presumably to prevent the dealer from gauging the buyer's level of interest. Similarly, in the realm of therapeutics, Rubin (1965) has suggested that a psychiatric interviewer might closely observe a patient's pupils for information about his or her feelings and interests. This advice implies that mere dilations and constrictions of the pupil can serve as information processors, and that there are physical responses that mirror the thought process of the human brain.

Pupil dilation can be affected by several factors such as light reflex, age, novelty, arousal or sexual preference, taste, anxiety, cognitive load/overload, gender, luminance, habituation, mental unrest, alcohol level, schizophrenia, etc., to name a few (Janisse, 1977). However, it is often difficult to state confidently that dilation may have occurred due to cognitive load alone, or anxiety alone, and so on. The mind reacts to an interplay of several factors which could affect the variation of pupil size. Nevertheless, studies have successfully demonstrated the correlation of pupil dilation with various attributes of human cognition and emotions.

Hess and Polt (1960) established the relation between pupil size and sexual preference or favorable events. They found that pupils register certain activities of the nervous system in addition to the effects of visual stimulation. They conducted an experiment where pictures of nude males, nude females, mother and baby, and baby and landscape were shown to a group of participants consisting of males and females. They observed a clear sexual dichotomy in terms of interest value to the pictures such that male participants' pupils dilated more in response to nude female pictures and less to mother and baby pictures as compared to the female participants. Pupil dilation in the case of cats reared as pets was also studied. It was noticed that, under constant light conditions, there were significant dilations in response to stimuli such as a familiar object of play and food. When food was not recognized, maximal dilation of the pupil did not occur until the scent of the food reached the animal.

The pupil diameter has also been related to the amount of material which is under active processing during a short term memory task. Kahneman and Beatty (1966) demonstrated pupil dilation as a measure of cognitive load such that, as the cognitive load increases, the dilation also increases. They employed digit span tasks and increased the number of digits that were to be reproduced as the experiment progressed. They observed that the dilation was the maximum during loading of information for recall, especially at the termination of the sentence. At the same time the pupil size decreased during the recall or the "unloading" phase.

Anxiety response can also affect the degree of pupil dilation. Johnson (1971) noted that arousal was associated with the anticipation of a consequence to the condition at hand. He designed an experiment similar to the digit-span task, in that subjects heard five words, one every two seconds, which they were to recall. Sharp dilation-constriction cycles were observed where subjects were asked to remember and repeat the words in anticipation of the recall, i.e., when the subjects would anticipate the recall of the words at specific times as instructed by the researcher, it was reflected in the dilation of the pupil.

As can be concluded from these studies, pupils constantly communicate the mental state of the person. One question that arises is whether the same behavior is observed in the case of an interaction between humans and computers as in the case when humans interact with humans through the medium of computers. During our interaction with computers, is there an implicit belief that makes us treat computers as humans? Researchers in the field of human computer interaction focus on creating more and more intuitive interfaces, where the interaction between humans and machines is smooth. They try to build machines that are more human like. Why is there such a paramount need to create machines that behave like humans? The interface can indeed go a long way in bridging the gap between people and machines. But can the interface affect our pupillary reactions to stimuli? Is there a significant difference in the pupil response when humans interact with humans as opposed to when they interact with machines? To address this question, pupil dilation was investigated as study participants interacted face-to-face and via an online chat conversation.

METHODS:

Nine subjects participated in the study, which was conducted in the Interaction Design Lab at the School of Information and Library Science at the University of North Carolina at Chapel Hill. The study was divided into two phases. The first phase consisted of an online chat conversation facilitated by AOL Messenger and the second phase consisted of a face-to-face conversation with the researcher. In both phases, pupil dilation was measured. The study methods are described in more detail below.

Participants:

Nine subjects, seven males and two females, participated in the study. The subjects were graduate students studying at The University of North Carolina at Chapel Hill. They were all familiar with using instant messaging systems. All participants were English speakers.

Eye-tracking:

The pupil information of the participants during the online chat conversation and the face-to-face conversation was recorded using an ASL 504 commercial eye-tracker (Applied Science Laboratory). It utilizes a camera that reconstructs a subject's eye position through the Pupil-Center and Corneal-Reflection method. A software application accompanying the system was used for the simultaneous acquisition of the subject's pupil movements. Only one eye tracker was employed during the study.

The camera in the eye tracker obtains pupil information through the infrared beam that is reflected off the pupil. An initial calibration is obtained when the subject fixates on nine distinct points on the desktop screen. The magnetic head tracker facilitates the automated tracking of the subject's pupil in event of movement, thereby not restricting the user to remain steady throughout the experiment. Ambient light settings were adjusted to ensure that the dilation was not due to light reflex.

For the online chat conversation, an application analysis of the instant messaging system was performed using the gaze tracker that automatically obtains pupil information for the application open on the desktop window. The application was loaded in the gaze tracker to obtain pupil response. Two NTSC monitors constantly displayed pupil response and the subject's desktop.

For the face-to-face conversation, no application was loaded and the eye tracker was calibrated to obtain the pupil dilation independent of the desktop settings.

Study Procedures:

The entire study was divided into two distinct phases; the first phase consisted of an online chat conversation with the researcher and the second phase consisted of a face-to-face conversation. The conversation with each participant consisted of a set of pre-defined questions which were time-stamped, i.e. the time at which each question was asked was noted in case of both the sessions. The questions were designed in a manner such that they would generate discussion of two distinct events which involved similar consequences. For example, assassination questions (see below) of two famous personalities, such as Martin Luther King and John F. Kennedy, involved the discussion of the year they were assassinated, the subject's knowledge about the matter and personal views about them. These two different sets of questions were based on similar scenarios. This was to ensure that the subject's anticipation of familiar questions did not affect the behavior of the pupil. A sample set of questions is shown as follows:

- Are you aware about the assassination of Martin Luther King?
- How did you learn about it?
- Do you remember the year he was assassinated?
- Do you happen to know the reason why it happened?
- Do you know who was incarcerated for his assassination?
- What are your personal thoughts on this?

An instant messaging account was already created and initialized on the subject's desktop to ensure dyadic conversation with the researcher. The application was loaded for analysis by the gaze tracker. All other applications were closed to ensure the pupil measurements for only the chat interface were obtained. The chat conversations were saved for future reference.

In the second phase, the researcher was seated facing the subject, such that the camera settings were not disturbed. The eye-tracker forms a vector with the head unit, the camera and the magnetic tracker to obtain a constant reading of the pupil. The face-to-face conversation was audio-taped for comparative analysis with the online chat conversation. The participants were again asked to fixate on the nine distinct points on the screen. Ambient light settings were re-adjusted to ensure minimal interference of light reflex. In this phase, no application was loaded for the gaze tracker and all other existing applications were closed. The questions were once again time-stamped for reference while listening to the recorded version.

The entire analysis was done using the ERICA (Eye-Gaze Response Interface Computer Aid) software which records pupil data and generates the pupil graph and data points or size of the pupil with respect to the X and Y axis against time. In the online chat conversation, the AOL application was loaded in the ERICA software which automatically captured the subject's pupil once the application was launched.

At the end of the two phases, subjects were asked to fill out a post-session questionnaire that helped gauge their familiarity with the topics and concentration during the conversations. At the same time, the subjects were also asked if they felt any difference on an individual basis during the two chat sessions.

Measurement and Data Analysis:

The eye tracker calibrated the position and size of the pupil. The magnetic head tracker monitored the subject's movements with respect to the camera of the eye-tracker to ensure accurate readings of the pupil during the study. There were a few problems with the calibration of the pupil, while the subject was looking at the keyboard during typing, in the online chat conversation. But the response was mainly noted for the time when the subject was asked the question and the gaze during the response from the researcher.

Data from two participants, out of the total nine, were eliminated because a good pupil reading was not obtained during the chat conversation as well as the face-to-face conversation.

The face-to-face conversation was audio-taped to note the time intervals at which questions were asked and the responses from the subjects were received. The time at which questions in the online-chat conversation were asked and the responses received were also monitored for a comparative analysis of the pupil behavior. The post-session questionnaire responses were compared to the behavior of the pupil during the study to observe any significant dilation or constriction with respect to the questions.

The ASL software for the eye tracker generates data points for the pupil dilation with respect to the X and Y coordinates at discrete time intervals. These time intervals were compared with the time-stamps maintained during the study, and the corresponding dilations were obtained. Dilations were recorded for each question in both phases and the mean was calculated.

A sample pupil graph generated by ERICA for one of the participant is shown in Figures 1 and 2 for face-to-face and online chat conversation, respectively. The dilations observed in both were compared and the mean, minimum and maximum dilations were plotted.

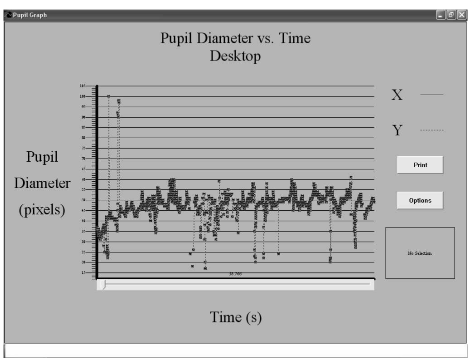

Figure 1: Pupil graph for face-to-face conversation

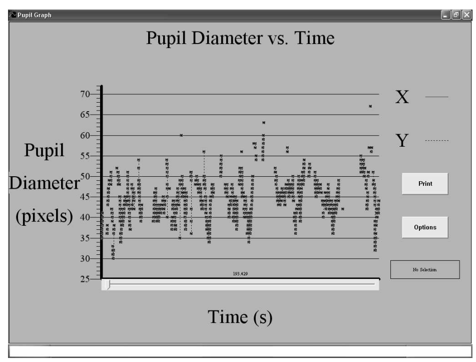

Figure 2: Pupil graph for online chat conversation

RESULTS:

Figure 3 indicates the mean dilation of the pupil of the participants for questions 1,2,3,4 and 5. It can be noted that the mean dilation was much higher in the face-to-face conversation as opposed to the online chat conversation. As discussed earlier, similar ambient light settings and questions were maintained in both the phases to ensure minimal noise due to light reflex and familiarity or anticipation of questions.

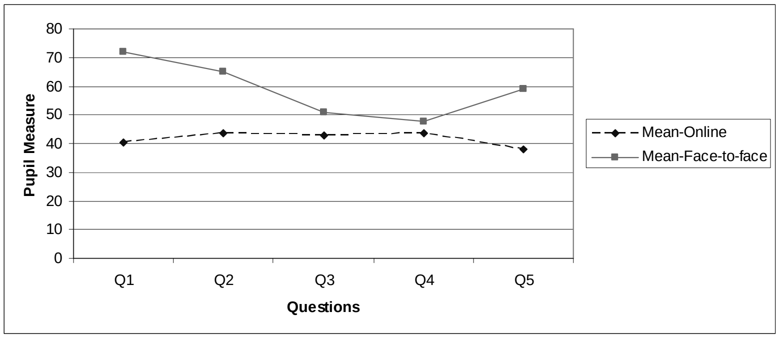

Figure 3: Mean pupil dilation for online chat v/s face-to-face conversation.

Figure 4 indicates the minimum and the maximum size of the pupil observed for each question. The maximum dilation during face-to-face conversation was higher than the maximum during online chat conversation. All the participants showed significant changes in pupil dilation during the face-to-face conversation.

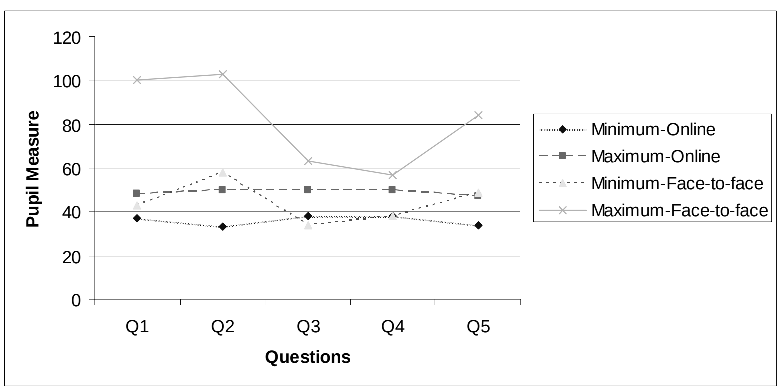

Figure 4: Minimum and Maximum Pupil Dilation

DISCUSSION:

The results discussed above were obtained only for a dyadic conversation in each of the phases. In an online chat conversation, a user usually encounters several distractions such as pop-up windows due to other running applications. Hence, it is often difficult to maintain undivided attention during such online chat sessions. Kiss et al. (2004) maintain that paying attention means that the guiding gestures are followed by an appropriate shift of gaze and attention to regions or objects that are involved in the conversation, which can affect our pupil behavior. For instance, in a case where two people are having a discussion whether to buy the table in the store or not, they often tend to point at the table and shift their gaze between the listener and the object of the discussion. The case could, however, differ in a computer mediated conversation.

Also, Argyle and Cook (1977) have found that, when two people talk, 60% of conversation involves gaze, 30% involves mutual gaze, and people look nearly twice as much (75%) while listening as compared to speaking (41%). This is seldom possible except in a dedicated chat conversation using an instant messaging system. However, at the time when the person responds (or virtually speaks) in an online chat conversation, the dilations should be approximately the same as in a face-to-face conversation, as again, similar thought processes of arousal, anxiety, fatigue or cognitive load would affect the changes in the pupil. But as the results indicate, it is not so. Dilations on questions 1, 2 and 5 were noticeably different than on questions 3 and 4, in the case of face-to-face conversation. These differences can be attributed to the effects of cognitive load and interest which can significantly affect the size of the pupil. Also, in a computer mediated conversation, it is often difficult to completely focus on the chat window, as the fundamental concept of turn-taking is absent, which makes it difficult to gauge the floor of the conversation. The level of interest generated in a face-to-face conversation, which is usually governed by our emotions, is also not easily created when using an instant messaging system.

Machines definitely cannot substitute for humans; they can, however, be modeled to exhibit human behavior. The whole course of interaction is governed by conversational factors of gaze, emotions and body language. These are indeed difficult to model with machines. Mere sense of acknowledgement through gaze, pupil response and body language gives reassurance to humans that their conversational partner is listening to them. This essential element is missing in the medium of machines. Instant messaging systems are good abstractions for human-human interaction. But as can be observed from the results, they are indeed not that engaging. Pupil dilations occur due to factors such as anxiety, arousal, interest, cognitive load, etc. The questions were designed so that these elements could affect the size of the pupils in the participants. However, the responses differed significantly in the two environments. When two people interact, they feel obligated to respond, whereas this is not necessarily required in case of interaction mediated by machines. This factor of "waiting to respond immediately" is not manifest in the case of messaging systems and, hence, people can take the liberty to feel distracted while participating in a conversation. There is no immediacy in the response since the level of interest between the "speaker-listener" roles is not known. Messaging systems also inculcate a sense of pre-occupation with other work that can affect the speaker's behavior in initiating a conversation. Our digital presence does not necessarily substitute our physical presence which is replete with engaging conversational feedback.

CONCLUSION:

Recent studies focus on designing interactive or attentive models for machines that will be capable of understanding the conversational cues employed by humans. But there still remains a need to bridge the gap between humans and machines, where humans can interact more freely with machines. There is plenty of evidence that mutual gaze is one of the conversational resources we rely on to successfully collaborate in conversation, but in the case of machines the demand for attention indeed depends on the motivation to perform work without any distractions. Often there is a certain amount of immediacy in response associated with a face-to-face conversation which is an instrumental factor not incumbent with machines. The human computer interaction could potentially be improved if computers were designed as human avatars. Video-chats are to an extent a good substitute, but the processing delays associated with them are enormous. Research should focus on designing more and more human-like interfaces where users are able to witness reality in a digital world. Pupil dilation is an important metric for studying human cognition and can lead us to a better understanding of human comprehension of systems. If similar dilations are observed, then it would certainly imply that machines are perceived as humans where they are able to generate the same level of cognitive activity as would otherwise be generated as a response to humans.

<u>REFERENCES:</u>

Argyle, M. & Cook, M. (1977). *Gaze and Mutual Gaze.* London: Cambridge University Press.

Gump, R. (1962). *Jade: Stone of heaven.* New York: Doubleday.

Hess, E. H., & Polt, J.M. (1960). Pupil size as related to the interest value of visual stimuli. *Science, 132,* 349-350.

Janisse,M.P. (1977). *Pupillometry: The Psychology of the Pupillary Response,* Washington, Hemisphere Publishing Corporation.

Johnson, D.A. (1971). Pupillary responses during a short term memory task: Cognitive processing, or arousal, or both? *Journal of Experimental Psychology, 90,* 311-318

Kahneman, D., & Beatty, J. (1966) Pupil diameter and load on memory. *Science, 154,* 1583-1585.

Kiss, B., Benedek, B., Szijarto, G., and Taka`cs, B.(2004). Close loop dialog model of face-to-face communication with a photo real virtual human. *Visual Communication and Image Processing 2004 (EI25), SPIE Electronic Imaging*

Knapp, M.L. & Hall, J.A (1997). *Nonverbal Communication in Human Interaction*. New York , Holt Rinehart and Winston.

Maglio, P., Matlock, T., Campbell, C. , Zhai, S., Smith, B., (2000). Gaze and speech in attentive user interfaces. *Proceedings of the Third International Conference on Multimodal Interfaces, Bejing, China.*

Oh, A., Fox, H., Kleek, M., Adler, A., Gajos, K., Morency, L. & Darell, T. (2002). Evaluating look-to-talk. In *Extended Abstracts of CHI 2002.* Minneapolis: ACM Press, 650-651.

Senders, J. W. (1983). *Visual Scanning Processes.* Netherlands: University of Tilburg Press.

Rubin, L. S. (1965). Autonomic dysfunction in neurotic behavior. *AMA Archives of General Psychiatry, 12,* 572-585.

Short, J., William, E. & Christies, B. (1976). *The Social Psychology of Telecommunications*. London: Wiley

Vertigaal, R., & Ding, Y.(2002). Explaining effects of eye gaze on mediated group conversations: Amount or synchronization? In *Proceedings of CSCW 2002*. New Orleans: ACM Press.

Vertigaal, R., Slagter, R., Veer, V., & Nijholt, A. (2000). Why conversational agents should catch the eye. In *Extended Abstracts of CHI'2000.* The Hague, The Netherlands: ACM, 257-258.

Zhai, S. (2003).What's in the eyes for attentive input. *Communications of the ACM, 46*(3), 34-39.